新集群部署：

后台还没布上，客户端发给后台一个地址，后台仍然返回客户端一个旧地址，所以新流媒体还待部署。。。

<http://1.119.138.66:10080/index.html> 心神学堂

<http://1.119.138.66:10082/dataLayout> 智慧教室

18401605753 13241283459

更改xinshen-sm-01的root密码为XS-root-123

sudo passwd root更改root密码命令。

Hadoop生态集群搭建

前提安装JDK，apt-get update

JDK安装目录/root/jdk1.8.0\_141

apt-get install default-jdk等，这里已经安装好了1.8版本的jdk。Java -version

查询java安装路径: update-alternatives --display java

1.12没找到，重新安装（出现问题apt-get -f install，未解决）

1.7 /usr/lib/jvm/java-8-oracle/jre/bin/java

1.8 /usr/lib/jvm/java-8-oracle/jre/bin/java

1.11 /usr/lib/jvm/java-8-oracle/jre/bin/java

设置SSH无密码登录

安装SSH，安装rsync

每一台都生成一个公钥，写在一起

生成密钥：ssh-keygen -t dsa -P '' -f ~/.ssh/id\_dsa

位置： /root/.ssh/id\_dsa. /root/.ssh/id\_dsa.pub.

将产生的公钥（key）放置到许可证文件中：cat ~/.ssh/id\_dsa.pub >> ~/.ssh/authorized\_keys

下载安装Hadoop

wget [http://mirror.bit.edu.cn/apache/hadoop/common/hadoop-3.1.1/hadoop-3.1.1.tar.gz && tar -zxvf hadoop-3.1.1.tar.gz](http://mirror.bit.edu.cn/apache/hadoop/common/hadoop-3.1.1/hadoop-3.1.1.tar.gz%20&&%20tar%20-zxvf%20hadoop-3.1.1.tar.gz)

移动目录：mv hadoop-3.1.1 /usr/local/hadoop

设置hadoop环境变量，在/.bashrc中：

vim ~/.bashrc 并添加如下内容：

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

export HADOOP\_HOME=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_HOME/bin

export PATH=$PATH:$HADOOP\_HOME/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export HADOOP\_YARN\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib"

export JAVA\_LIBRARY\_PATH=$HADOOP\_HOME/lib/native:$JAVA\_LIBRARY\_PATH

source ~/.bashrc

修改hadoop配置文件，五个，

Hadoop-env.sh（hadoop的配置文件，必须设置java的安装路径）

core-site.xml（在这儿设置HDFS的默认名称）

yarn-site/xml（含有Mapreduce2{yarn}的相关配置）

mapred-site.xml（设置监控Map和Reduce程序的jobTracker任务分配情况以及TaskTracker的任务运行情况）

hdfs-site.xml（用于设置HDFS分布式文件系统）

vim /usr/local/hadoop/etc/hadoop/hadoop-env.sh 并添加：

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

vim /usr/local/hadoop/etc/hadoop/core-site.xml

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

vim /usr/local/hadoop/etc/hadoop/yarn-site.xml

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

复制模板文件(我在这里做了个备份)，由mapred-site.xml.template至mapred-site.xml

cp /usr/local/hadoop/etc/hadoop/mapred-site.xml /usr/local/hadoop/etc/hadoop/mapred-site.xml.template

设置mapreduce的框架为yarn：

vim /usr/local/hadoop/etc/hadoop/mapred-site.xml

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

设置namenode和datanode的数据存储目录：

vim /usr/local/hadoop/etc/hadoop//hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<peoperty>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop/hadoop\_data/hdfs/namenode</value>

</property>

<peoperty>

<name>dfs.datanode.data.dir</name>

<value>file:/usr/local/hadoop/hadoop\_data/hdfs/datanode</value>

</property>

接着创建刚刚在配置文件中设置的目录，哈哈，并将hadoop目录的所有者更改为xinshen（因为linux是多人多任务的操作系统，所有目录或文件都具有所有者，chown可以将目录或文件所有者更改）？

sudo mkdir -p /usr/local/hadoop/hadoop\_data/hdfs/namenode

sudo mkdir -p /usr/local/hadoop/hadoop\_data/hdfs/datanode

sudo chown xinshen:xinshen -R /usr/local/Hadoop

HDFS进行格式化

hadoop namenode -format

启动Hadoop

1. 启动HDFS:start-dfs.sh
2. 启动YARN：start-yarn.sh

3.也可以用start-all.sh

4.jps(集群未启动，单节点)

1. 修改host文件

vim /etc/hosts

192.168.1.7 master-hadoop

192.168.1.8 slave2-hadoop

192.168.1.11 slave3-hadoop

1. 配置slaves

关键技术解释：

Pig：基于Hadoop的大规模数据分析平台，pig为复杂的海量数据并行计算提供简单的操作和编程接口。

Hdfs：基于hadoop的工具，提供完整的SQL查询，可将sql语句转换为Mapreduce任务进行运行。

Mapreduce：一种编程模型，用于大规模数据的的并行计算

Yarn

Shuffle

Hive